



Shiromani Gurdwara Parbandhak Committee's

**Guru Nanak Khalsa College of Arts, Science & Commerce
(Autonomous)
Matunga,
Mumbai – 400 019,
Maharashtra**

Program: Master of Science

Course: Bioinformatics

Syllabus for M.Sc. Semester I & II

**(As per Choice Based Semester and Grading System with effect
from Academic Year 2023-2024)**

Introduction:

Last two decades have witnessed the emergence of Bioinformatics as a major thrust area in the global scientific scenario. It is the interface between the two most rapidly advancing fields of biological and computational sciences. Bioinformatics is a scientific discipline and asset of skills that has now become one of the most important information gathering, data mining, and knowledge building tools in recent research and clinical applications.

In the current pandemic bioinformatics has played a major role in studying the genomics of COVID-19. It helped in studying the impact of Hydroxychloroquine on COVID. Bioinformatics came out as one of the essential tools to analyze viral data as it provides vital information about the genetic makeup of the virus and assists directly in the development of drugs or vaccines against the deadly disease.

This course is of two years imparting technical skills for producing industry ready candidates. Post Graduate Bioinformatics course content includes Biological sciences (Cell biology, Biochemistry, Molecular biology, Proteomics, Immunology), Biostatistics, Machine learning, Chemical sciences (Drug designing, Analytical techniques like HPLC, GC, HPTLC, IR, Mass Spectroscopy), Computational sciences (Java, Linux, Perl, Mongodb, Python), Clinical Research and Clinical Data Management, Bioinformatics in Genomics and Proteomics, Sequence Analysis, Structural Biology, Drug Designing, Molecular Modeling and Simulations, Cheminformatics, Immunoinformatic, QSAR, Research Methodology and Scientific Writing.

In the second-year students are introduced to Research Methodology and Scientific Writing. Research culture is inculcated by pursuing a compulsory 3 months dissertation project. They are trained to present their research work in scientific conferences and research meet. They also publish their research work in eminent journals. This adds a feather to cap while they apply for PhD programs in India and abroad.

The course is specially designed to cater to the requirements of the pharmaceutical and biotech industries as well as emerging sectors like biopharmaceuticals, clinical research, data science, and information technology. Opportunities are arising in areas like big data, data analytics, drug designing, bio-analytics, hospitals, research & development labs, data Science companies, biotechnology & healthcare companies, sequence assembly, genomic sequence analysis, functional genomics, genotyping, pharmacogenomics, database administration, etc.

We are blessed with a team of renowned scientists and faculty members across industry and academics who are prompt and enthusiastic and have helped students build lucrative careers in the said fields.

The proposed program for Masters in Bioinformatics is a two-year program which has been successful in developing young minds into bright researchers and industry professionals.

Objectives of The Course:

- Develop trained manpower in the field of Bioinformatics with specific emphasis for fulfilling expectations of Indian Pharma and Biotech industry.
- To develop skills needed to collect, understand, analyze and manage data generated through high throughput technology.
- Explain how the boundaries of knowledge in this professional discipline are advanced through research and enable students to conduct research through targeted research projects.
- To give exposure to the students in national and international work.

Eligibility:

Bachelor degree from a recognized university in any of the following disciplines:

- Science (Biochemistry, Biology, Bioanalytical Science five years integrated, Biophysics, Botany, Biotechnology, Chemistry, Environmental Science, Life Sciences, Physics, Mathematics, Microbiology, Statistics or Zoology)
- Agriculture
- Medicine
- Pharmacy
- Veterinary Science
- Computer Science

Staffing Pattern for The Course:

- Instrument technician - 01
- Technical assistance - 02
- Lecturer - 03 (Full Time)
- Lecturer - 01 (Part Time)
- Remaining workload will be shared with visiting faculty.

Faculty:

Post graduate or PhD degree in the subject of Bioinformatics, Bioanalytical Science, Biochemistry, Biotechnology, Chemistry, Computers Science, Information technology, Microbiology, Life Science, Molecular Biology, Botany, Zoology, and Statistics, Mathematics with B+ grade and NET / SET.

SYLLABUS IN BRIEF

Distribution of Credits

Semester I

Paper code	Paper	Paper Name	Credit
GNKPSBI1501	Mandatory Paper - I	Fundamental of Biology & Bioinformatics	4
GNKPSBI2501	Mandatory Paper- II	Introduction to Programming Languages and Databases	4
GNKPSBI3501	Elective Paper	Bioinformatics and Sequence Analysis	3
GNKPSBI4501	Research Methodology	Research Methodology	2+2 = 4
GNKPSBI1P501	Mandatory Paper I – Practical	Fundamental of Biology & Bioinformatics	2
GNKPSBI2P501	Mandatory Paper II – Practical	Introduction to Programming Languages and Databases	2
GNKPSBI3P501	Elective Paper - Practical	Bioinformatics and Sequence Analysis	1

Semester II

Paper code	Paper	Paper Name	Credit
GNKPSBI1502	Mandatory Paper-I	Advance Biology & Structural Bioinformatics	4
GNKPSBI2502	Mandatory Paper-II	Java Programming, Introduction to Linux and Machine Learning	4
GNKPSBI3502	Elective Paper	Genomics & Proteomics in Bioinformatics, NGS	3
GNKPSBI4502	ON JOB TRAINING	ON JOB TRAINING	4
GNKPSBI1P502	Mandatory Paper I – Practical	Advance Biology & Structural Bioinformatics	2
GNKPSBI2P502	Mandatory Paper II – Practical	Java Programming, Introduction to Linux and Machine Learning	2
GNKPSBI3P502	Elective Paper - Practical	Genomics & Proteomics in Bioinformatics, NGS	1



Shiromani Gurdwara Parbandhak Committee's

**Guru Nanak Khalsa College of Arts, Science and Commerce
(Autonomous)
Matunga, Mumbai – 400 019, Maharashtra**

Syllabus for M.Sc. Semester I

Program: Master of Science

Course: Bioinformatics

Mandatory Paper: I

**(As per Choice Based Semester and Grading System with effect
from Academic Year 2023-2024)**

M.Sc. Bioinformatics (2023-

2024) Mandatory Paper I

Course Title: Fundamental of Biology & Bioinformatics

Paper Code: GNKPSBI1501

Credits: 04

No. of Lectures:

60 Marks: 100

Learning Objectives:

- To learn about the bioinformatics databases, databanks, data format and data retrieval from the online sources.
- To make students understand the essential features of the interdisciplinary field of science for better understanding biological data.
- To provide the student with a strong foundation for performing further research in bioinformatics.
- To understand and explore the sequence comparison tools for nucleotide and protein sequences.
- To interpret the results of sequence alignments (Pairwise - local and global alignments as well as multiple sequence alignments).
- To study basis of cell biology.
- This course will enable students to understand structure of Biomolecules and their Metabolism.
- To lay a foundation of Enzyme Kinetics.

Learning Outcome:

- Gain an overview of cells - their origin and evolution. Students will understand the structures and purposes of basic components of prokaryotic and eukaryotic cells, especially macromolecules, membranes, and organelles. They will apply their knowledge of cell biology to selected examples of changes or losses in cell function.
- Students will be able explain/describe the synthesis of proteins, lipids, nucleic acids, and carbohydrates and their role in metabolic pathways along with their regulation. Students will be able to use current biochemical techniques to plan and carry out experiments.
- Introduce what bioinformatics is and why it is important.
- Provide an overview of the application areas of bioinformatics, with a focus on the topics that will be taught.
- Locate and use the main databases at the NCBI and EBI resources.
- Know the difference between databases, tools, repositories and be able to use each one to extract specific information.

Semester I

Unit		Title	No. of lectures
Unit 1		Cell Biology	15
	1.1	Structure and Function of cells and intracellular organelles (of both prokaryotes and eukaryotes)	
	1.2	Cell Membranes- fluid mosaic model and its functions. Membrane transport, Na-K Pump	
	1.3	Structure, various models, its function. Cellular transport - simple diffusion, facilitated diffusion, active transport, nuclear transport, osmosis exocytosis and endocytosis	
	1.4	Structure and Function of cells and intracellular organelles (of both prokaryotes and eukaryotes)	
	1.5	Cytoskeleton- Introduction to microfilaments, microtubules, and intermediate filaments, Nucleation	
	1.6	Mechanism of cell division including (mitosis and meiosis) and concept of motors. Programmed cell death	
	1.7	Characteristics of cancerous cells; Agents promoting carcinogenesis; molecular basis of cancer therapy, Tumor markers - AFP, CEA, hCG; Telomere replication; Telomerase and its role in cancer and aging.	
	1.8	Neurotransmitters	
Unit 2		Structure of Biomolecules and their Metabolism	15
	2.1	Nucleic Acids: RNA and DNA	
	2.2	Different Structural Forms of DNA; Different Types of RNA.	
	2.3	Carbohydrates: Structure and Classification, Glycosaminoglycans, Proteoglycans and Glycoproteins	
	2.4	Lipids: Structure and Classification, Acylglycerols; Phospholipids, Sphingolipids, Cholesterol and Membranes; Isoprenoids, Icosanoids and their biological importance	
	2.5	Carbohydrates: Structure and Classification, Glycosaminoglycans, Proteoglycans and Glycoproteins	
	2.6	Glycolysis, Feeder, pyruvate oxidation, the Krebs cycle, and oxidative phosphorylation.	
Unit 3		Enzyme Kinetics	15
	3.1	Introduction to Enzymes	
	3.2	Units of activity, coenzymes and metal cofactors.	
	3.3	Temperature and PH effects	
	3.4	Michaelis – Menten Kinetics	
	3.5	Feedback inhibition and activation	
	3.6	Ribozyme and abzymes	

Unit 4		Introduction to Bioinformatics & Sequence Analysis	
	4.1	Introduction to the Bioinformatics and its application	15
	4.2	Importance, goals / scope, opportunities & Limitations of Bioinformatics	
	4.3	Various file formats for bio- molecular sequences	
	4.4	Basic Concepts of sequences similarity, Identity, homology, orthologues, Paralogues	
	4.5	Concepts of sequence alignments, Dot Matrix, Dynamic Programming, Heuristic Programming,	
	4.6	Algorithms: Needleman & Wunch, Smith & Waterman	
	4.7	Scoring matrices: Basic concepts of scoring matrix, PAM and BLOSUM series and principles based on which these matrices are derived	

Semester I Evaluation Pattern

Internal: 25

External: 75

M.Sc. Bioinformatics (2023-2024)

Mandatory Paper I - Practical

Course Title: Fundamental of Biology & Bioinformatics

Paper Code: GNKPSBI1P501

Credits: 02

Marks: 50

1. Biochemical estimation of DNA.
2. Biochemical estimation of RNA.
3. Estimation of vitamin C using UV- Vis spectrophotometer.
4. 2D separation of amino acids, plant pigments using paper chromatography.
5. Thin layer chromatography analysis of modern drug.
6. Pair-wise global alignments of protein/ DNA sequences using Needleman-Wunsch algorithm
7. Pair-wise local alignments of protein and DNA sequences using Smith-Waterman algorithm
8. Identification of similar and divergent sequences:
 - a. BLAST
 - b. FASTA
 - c. PSI BLAST
 - d. PHI BLAST

References:

- Bioinformatics: Sequence and Genome Analysis by Mount D., Cold Spring Harbor Laboratory Press, New York. 2004
- Bioinformatics- a Practical Guide to the Analysis of Genes and Proteins by Baxevanis, A.D. and Francis Ouellette, B.F., Wiley India Pvt Ltd. 2009
- Introduction to Bioinformatics by Teresa K. Attwood, David J. Parry-Smith. Pearson Education. 1999
- Introduction to Bioinformatics Algorithms, Neil C. Jones, PavelPevzner
- Introduction to Bioinformatics - S.Sundararajan and Balaji
- Instant notes – Bioinformatics – Westhead, Howard parish and Twyman. Viva books Pvt. Limited. Chennai
- Bioinformatics basic methods and applications – S.C. Rastogi
- Essential Bioinformatics by Jin Xiong, 2006
- ORENGO, C., JONES, D. & THORNTON, J.: Bioinformatics: genes, proteins and computers. 2003. Bios Scientific Publishers, Ltd. Oxford
- Ingvar Eidhammer, Inge Jonassen, William R. Taylor: Protein Bioinformatics. 2003 John Wiley And Sons Ltd (UK) 2. HIGGINS, D. & TAYLOR, W.: Bioinformatics: sequence, structure, and databank. 2000. Oxford University Press, Oxford, UK
- M. Lesk. (2002) Introduction to Bioinformatics Oxford University Press 68. G.B Fogel D.W Corne (2002) Evolutionary Computation in Bioinformatics
- Voet, Donald, Voe Judith, Pratt, Charlotte W. Fundamentals of Biochemistry: Life at the molecular Level 2nd Edition. Publisher: Asia, John Wiley & Sons. 2006.
- Nelson David L., Cox Michale. Lehninger Principles of Biochemistry 5th Edition.Publisher: New York. W. H. Freeman. 2008.
- R C Rastogi, Biochemistry 2nd edition, 2003.
- Kleinsmith and Harden, The World of the cell, Becker, Academic Internet Publishers; 5th edition (2006)
- Geoffrey M. Cooper and Robert E. Hausman. The Cell: A Molecular Approach, Fourth Edition
- Harvey Lodish. Molecular cell Biology. W. H. Freeman; Sol edition (2007)



Shiromani Gurdwara Parbandhak Committee's

**Guru Nanak Khalsa College of Arts, Science and Commerce
(Autonomous)
Matunga, Mumbai – 400 019, Maharashtra**

Syllabus for M.Sc. Semester I

Program: Master of Science

Course: Bioinformatics

Mandatory Paper: II

**(As per Choice Based Semester and Grading System with effect from
Academic Year 2023-2024)**

M.Sc. Bioinformatics (2023- 2024)

Mandatory Paper II

Course Title: Introduction to Programming Languages and Databases

Paper Code: GNKPSBI2501

Credits: 04

No. of Lectures:60

Marks: 100

Learning Objectives:

- Explore various programming paradigms as well as principles of building object-oriented software
- Solidify your understanding of Computer Science in theory and practice related to biological data.
- Learn the essential elements of computing theory including logic, proof techniques, combinatorics, algorithm analysis, discrete data models (sets, relations, trees), and finite automata.

Learning Outcomes:

- Understand dynamic memory management techniques using pointers, constructors, destructors, etc.
- Classify inheritance with the understanding of early and late binding, usage of exception handling, generic programming.
- Understand generic programming, templates, file handling.

Unit		Title	No. of lectures
Unit 1		Fundamentals of Computing and introduction to Database systems	15
	1.1	Overview and functions of a Computer System	
	1.2	Memory: Register, buffer, RAM, PROM, EPROM, EEPROM	
	1.3	Overview of various computer threats viz. Trojan horses, viruses etc.	
	1.4	Operating systems concepts (Windows, Unix, Linux)	
	1.5	Software Development Life cycle (SDLC)	
	1.6	Concepts of various types of Databases and Data models	
	1.7	Instances, Schemes and E – R Model	

Unit 2		Introduction to C++	15
	2.1	Input and Output	
	2.2	Data types and Functions	
	2.3	Derived data types	
	2.4	Structure in C++ and string manipulations	
	2.5	Basic Input output and File handling	
Unit 3		Advanced C++	15
	3.1	Basics of object-oriented programming	
	3.2	Concept of Object, class, Constructor and Destructors	
	3.3	Data encapsulation	
	3.4	Inheritance	
	3.5	Polymorphism	
	3.6	Virtual function and friend function	
Unit 4		SQL	15
	4.1	Data definition statement; Data Manipulation Statements	
	4.2	Select statement	
	4.3	Data Control Statement	
	4.4	Other Database Objects (Views, Sequences, Synonyms etc)	
	4.5	Security, Grants, Roles, Privileges	

Semester I Evaluation Pattern

Internal: 25

External: 75

M.Sc. Bioinformatics (2023-2024)

Mandatory Paper II - Practical

Course Title: Introduction to Programming Languages and Databases

Paper Code: GNKPSBI2P501

Credits: 02

Marks: 50

1. Basic I/O Programs
 - a. cin
 - b. cout
 - c. Datatypes and variables
2. Conditional and Loops
 - a. If statement

- b. If-else statement
 - c. Nested If-else statement
- 3. Arrays
 - a. 1-D array
 - b. 2-D array
- 4. Functions and structure
 - a. Create functions with no return and with no parameter
 - b. Create functions with return and with parameter
 - c. Create functions with no return and with parameter
 - d. Create functions with return and with no parameter
- 5. String manipulation
 - a. strlen()
 - b. strcpy()
 - c. strcat()
- 6. Class and objects, Encapsulations
 - a. Create class and objects
 - b. Access specifiers: Private, public and protected
- 7. Inheritance
 - a. Single Inheritance
 - b. Multilevel Inheritance
 - c. Multiple Inheritance
 - d. Hybrid Inheritance
- 8. Polymorphism, Virtual Function, Friend Function
 - a. Runtime Polymorphism
 - b. Compile Polymorphism
- 9. Constructor and Destructor
- 10. SQL
 - a. Create table
 - b. Insert values
 - c. Add new column to table
 - d. Remove the record from table
 - e. Constraints
 - f. Calculate average, sum and max
 - g. group by, where, order by

References:

- 1. C++: The Complete Reference, 4th Edition
 - 2. Advanced C++ by BPB.
 - 3. SQL The Complete Reference, 3rd Edition
-



Shiromani Gurdwara Parbandhak Committee's

**Guru Nanak Khalsa College of Arts, Science and Commerce
(Autonomous)
Matunga, Mumbai – 400 019, Maharashtra**

Syllabus for M.Sc. Semester I

Program: Master of Science

Course: Bioinformatics

ELECTIVE PAPER

**(As per Choice Based Semester and Grading System with effect from
Academic Year 2023-2024)**

M.Sc. Bioinformatics (2023-2024)

Elective Paper

Course Title: Bioinformatics and Sequence Analysis

Paper Code: GNKPSBI3501

Credits: 03

No. of Lectures: 45

Marks: 100

Learning Objectives:

- To learn about the bioinformatics databases, databanks, data format and data retrieval from the online sources.
- To make students understand the essential features of the interdisciplinary field of science for better understanding biological data.
- To provide the student with a strong foundation for performing further research in bioinformatics.
- To understand and explore the derived databases and their applications to biological systems.
- To explore various computational tools for analysis of biomolecular sequences.

Learning Outcomes:

- Provide an introduction to what bioinformatics is and why it is important.
- Provide an overview of the application areas of bioinformatics, with a focus on the topics that will be taught.
- Explain what type of knowledge will be gained.
- Locate and use the main databases at the NCBI and EBI resources.
- Know the difference between databases, tools, repositories and be able to use each one to extract specific information.
- Define and Identify different life skills required in personal and professional life
- Develop an awareness of the self and apply well-defined techniques to cope with emotions and stress.
- Explain the basic mechanics of effective communication and demonstrate these through presentations.
- Take part in group discussions CO5: Use appropriate thinking and problem solving techniques to solve new problems
- Understand the basics of teamwork and leadership

Unit		Title	No. of lectures
Unit 1		Biomolecular Databases	15
	1.1	What are Database, Types of Databases & Biological Databases	
	1.2	Major Bioinformatics resources NCBI, ENTREZ, EMBL-EBI & ExPasy	
	1.3	Literature database e.g. PUBMED & PMC	
	1.4	Nucleic acid database: GENBANK, EMBL, DDBJ Study of Submission Tools	
	1.5	Protein sequence databases: PIR, UniProt, SWISSPROT, TrEMBL, PROSITE, Pfam, INTERPRO / PRINTS	
	1.6	Protein structure databases: NDB, PDB, CCSD, Reactome, PDBSum, PDBTM, CATH, SCOPe	
	1.7	Genome Databases: NCBI, UCSC, GOLD, Ensemble,	
	1.8	Chemical Databases: PubChem, Drug Bank, ZINC	
	1.9	Specialized & Metabolic databases: MBGD, ICTVdb, KEGG, OMIM	
	1.10	Enzyme Database: BRENDA, EXPASY	
Unit 2		Molecular Phylogenetics	15
	2.1	Introduction to Multiple sequence alignment(MSA) and its various algorithm.	
	2.2	Introduction to Multiple sequence alignment(MSA) and its various algorithm.	
	2.3	Definition and description of Phylogenetic trees, Basics & Nature of data in Taxonomy and Phylogeny.	
	2.4	Concept of evolutionary trees – Dendrogram and Phylogram, its representations & Procedure for Phylogenetic tree construction	
	2.5	Methods for Tree construction:	

		a) Distance based method: Clustering based method (NJ & UPGMA), Optimality based method (FM and ME) b) Character based method: Maximum Parsimony and Maximum likelihood method	
	2.6	Phylogenetic Analysis Programs: PHYLIP, PAUP, MEGA4	
Unit 3		OMICS	
	3.1	Introduction to Genomics, Genome annotation and analysis, DNA sequencing technologies, Comparative genomics, Structural genomics, Functional genomics, Pharmacogenomics, Personalized medicine and genomics	15
	3.2	Introduction to Proteomics, Protein separation and purification techniques, Mass spectrometry and proteomic analysis, Protein identification and characterization, Protein-protein interactions, Protein expression profiling	
	3.3	Introduction to Epigenomics & Structural Biology	
	3.4	Goals of the Human Genome Project, cloning vectors, concept of maps, physical maps, shotgun libraries, DNA polymorphism, nucleotides, DNA sequences.	
	3.5	Restriction enzymes, resource for restriction enzyme (REBASE)	
	3.6	Bioinformatics & Computational Biology a. Data analysis methods for omics data b. Database resources for genomics, proteomics, and other omics data c. Statistical analysis and data mining (Application / case study) d. Machine learning and predictive modeling (Case study)	

Semester I Evaluation Pattern

Internal: 25

External: 75

M.Sc. Bioinformatics (2023-2024)

Elective Paper - Practical

Course Title: Bioinformatics and Sequence Analysis

Paper Code: GNKPSBI3P501

Credits: 01

Marks: 50

1. Exploring the integrated database system viz. Basic, Advanced and LIMITS w.r.t NCBI, EMBL-EBI, ExPasy
2. Sequence databases:
 - a. EMBL
 - b. DDBJ
 - c. GenBank & Submission Tool
 - d. UniProt
 - e. PIR
 - f. SWISSPROT & TrEMBL
3. Domain database:
 - a. PROSITE
 - b. INTERPRO / PRINT
 - c. Pfam
 - d. BLOCK
4. Structure database:
 - a. PDB
 - b. NDB
 - c. CCSD
 - d. Reactome
 - e. PDBSum
 - f. PDBTM
 - g. CATH
 - h. SCOPe
5. Literature & Specialized database:
 - a. KEGG
 - b. PUBMED
 - c. OMIM
 - d. PMC
6. Retrieving genomic information:
 - a. NCBI
 - b. GOLD & Submission Tool
 - c. Ensemble
 - d. UCSC
 - e. MGD
 - f. IGVdb

7. Multiple sequence alignment
 - a. Clustal Omega
 - b. T-Coffee
 - c. MUSCLE
8. Phylogenetic tree generation tool & Visualization:
 - a. Simple Phylogeny (EMBL-EBI)
 - b. Tree construction using PAUP & PHYLIP
 - c. Visualization of phylogenetic trees using Tree View
9. Restriction enzyme database: REBASE
10. Database resources for genomics, proteomics, and other omics data studies

References:

- Bioinformatics: Sequence and Genome Analysis by Mount D., Cold Spring Harbor Laboratory Press, New York. 2004
- Bioinformatics- a Practical Guide to the Analysis of Genes and Proteins by Baxevanis, A.D. and Francis Ouellette, B.F., Wiley India Pvt Ltd. 2009
- Introduction to Bioinformatics by Teresa K. Attwood, David J. Parry-Smith. Pearson Education. 1999
- Introduction to Bioinformatics Algorithms, Neil C. Jones, Pavel Pevzner
- Xinkun Wang, Next-Generation Sequencing Data Analysis, CRC Press, 2016, 1e 2. Publications on NGS data analysis from literature.
- G.P. Quinn and M. J. Keough; Experimental design and data analysis for biologists; Cambridge University Press, 2002.
- S. Knudsen; Guide to analysis of DNA microarray data; Wiley, 2nd edition, 2004.
- E. D. Hoffmann and V. Stroobant; Mass spectrometry: Principles and applications; Wiley, 3rd edition, 2007.
- A. Zhang; Advanced analysis of gene expression microarray data; World Scientific Publishing, 2006.
- N. Rodriguez-Ezpeleta; Bioinformatics for High Throughput Sequencing; Springer; 2012.
- Y. M. Kwon and S. C. Rieke; High-Throughput Next Generation Sequencing: Methods and Applications; Humana Press; 2011.
- J. R. Chapman; Protein and Peptide Analysis by Mass Spectrometry, Springer, 2nd edition, 1996.
- Introduction to Bioinformatics - S.Sundararajan and Balaji
- Instant notes – Bioinformatics – Westhead, Howard parish and Twyman. Viva books Pvt. Limited. Chennai
- Bioinformatics basic methods and applications – S.C. Rastogi
- Essential Bioinformatics by Jin Xiong, 2006
- ORENKO, C., JONES, D. & THORNTON, J.: Bioinformatics: genes, proteins and

computers. 2003. Bios Scientific Publishers, Ltd. Oxford

- Ingvar Eidhammer, Inge Jonassen, William R. Taylor: Protein Bioinformatics. 2003 John Wiley And Sons Ltd (UK) 2. HIGGINS, D. & TAYLOR, W.: Bioinformatics : sequence, structure, and databank. 2000. Oxford University Press, Oxford, UK
- M. Lesk. (2002) Introduction to Bioinformatics Oxford University Press 68. G.B Fogel D.W Corne (2002) Evolutionary Computation in Bioinformatics



Shiromani Gurdwara Parbandhak Committee's

**Guru Nanak Khalsa College of Arts, Science and Commerce
(Autonomous)
Matunga, Mumbai – 400 019, Maharashtra**

Syllabus for M.Sc. Semester I

Program: Master of Science

Course: Bioinformatics

RESEARCH METHODOLOGY

**(As per Choice Based Semester and Grading System with effect from
Academic Year 2023-2024)**

M.Sc. Bioinformatics (2023-2024)

Course Title: Research Methodology

Paper Code: GNKPSBI4501

Credits: 04 (02 + 02)

No. of Lectures: 60

Marks: 100

Learning Objectives:

- To be able to restate an investigative question in terms of a statistical model or algorithm
- Successfully relate theoretical concepts to a real-world problem in a written report.
- Provide reproducible statistical analyses using accepted practices of the research community
- Demonstrate the ability to find appropriate research literature appropriate to the investigative task.
- To understand the scope of personality and its development.
- To develop core skills for development of self.
- To cultivate interpersonal skills for successful life.

Learning Outcome:

- They have the versatility to work effectively in a broad range of analytic, scientific, government, financial, health, technical and other positions.
- They have a broad background in Mathematics and Statistics, an appreciation of how its various sub-disciplines are related, the ability to use techniques from different areas, and an in-depth knowledge about topics chosen from those offered through the department.
- They recognize and appreciate the connections between theory and applications
- Define and Identify different life skills required in personal and professional life
- Develop an awareness of the self and apply well-defined techniques to cope with emotions and stress.
- Explain the basic mechanics of effective communication and demonstrate these through presentations.
- Take part in group discussions CO5: Use appropriate thinking and problem-solving techniques to solve new problems
- Understand the basics of teamwork and leadership

Units	Topic	Lectures	Credits
1.1	Research Fundamentals and Terminology	07	01
	<p>Meaning and Objective of research, features of a good research study, types of Research (qualitative and quantitative research)</p> <p>Study designs and variations: basic, applied, historical, exploratory, experimental, ex-post-facto, case study, diagnostic research, crossover design, case control design, cohort study design, multifactorial design</p>		
1.2	Literature Survey Methods	08	01
	<p>Journal and abbreviation, current titles and review, monographs, textbooks, introduction to abstract, Beilstein, subject and author index</p> <p>Digital: Web sources, E-journals, Journal access, TOC alerts, Hot articles, Citation Index, Impact factor, Hindex, E-consortium, UGC infonet, E-books, Internet discussion groups and communities, Blogs, preprint servers, Search engines, Scirus, Google Scholar, Wiki-databases, Science Direct, SciFinder, Scopus.</p>		
2.1	Research Writing	10	01
	<p>Scientific writing- Reporting practical and project work, writing literature surveys and reviews, organizing a poster display, giving an oral presentation.</p> <p>Writing Scientific Papers: Justification for scientific contributions, bibliography, description of methods, conclusions, the need for illustration, style, publications of scientific work.</p> <p>Project Proposal and research funding agencies, Research grants, scholarships and funding (CSIR, DBT, DST, DST- INSPIRE Fellowship, ICMR, INSA, BRNS, MoEFCC, UGC-RFSMS, Fulbright)</p> <p>Fellowships for Indian students, Lady Tata Memorial Trust, EPA, Bill and Melinda Gates Foundation, Wellcome Trust, Erasmus Mundus)</p>		

2.2	Publication Ethics & Bibliography	05	
	<p>Publication ethics : definition, introduction</p> <p>Best practices/ Standards settings initiative and guidelines COPE, WAME</p> <p>Conflict of interest</p> <p>Publication Misconduct: definition, concept, problems that lead to unethical behavior</p> <p>Violation of publication ethics, authorship and contributorship</p> <p>Identification of publication misconduct</p> <p>Predatory publisher and journals. Use of reference management software (MS Word / Zotero / Mendeley)</p>		
3.1	Statistical Measures & Statistical Inference		
	<p>Mean, mode, median, Variance, standard deviation, coefficient of variation, Raw and Central Moments upto fourth order, measures of skewness and kurtosis</p> <p>Correlation and regression analysis, curve fitting (linear, non-linear, multilinear).</p> <p>Hypothesis testing: Simple and composite hypotheses, Null and alternative hypotheses critical region Type I and Type II Errors Level of significance p- value, power of a test</p> <p>Test of significance viz. Z test, t test, pair t test, chi2 test of goodness of fit</p> <p>Non-parametric tests: median test, sign test, Kruscal-Wallis test</p>	15	01

4.1	Communication Skills & Public Speaking		
	<p>Introduction to Personality Development The concept of personality, SWOT analysis, Attitude & Motivation Attitude - Concept - Significance - Factors affecting attitudes - Positive attitude – Advantages –Negative attitude- Disadvantages, Concept of motivation</p> <p>Self-esteem Term self-esteem - Symptoms - Advantages - Do's and Don'ts to develop positive self-esteem – Low self esteem - Symptoms - Personality having low self esteem - Positive and negative self esteem. Interpersonal Relationships – Defining the difference between aggressive, submissive and assertive behaviors - Lateral thinking</p> <p>Aspects of Personality Development Body language - Problem-solving - Conflict and Stress Management - Decision-making skills - Leadership and qualities of a successful leader – Character building -Team-work – Time management - Work ethics –Good manners and etiquette.</p> <p>The art of participating in Group Discussion – Facing the Personal (HR & Technical) Interview -Frequently Asked Questions - Psychometric Analysis - Mock Interview Sessions.</p>	15	01

Internal Evaluation: 50 Marks

External Evaluation: 50 Mark

References:

- BIOSTATISTICS by P.N. Arora & P.K. Malhan
- Biostatistics, B ANNADURAI
- Intuitive Biostatistics: A Nonmathematical Guide
- The Art of Public Speaking, STEPHEN LUCAS
- TED Talks: The Official TED Guide to Public Speaking, CHRIS ANDERSON



Shiromani Gurdwara Parbandhak Committee's

**Guru Nanak Khalsa College of Arts, Science and Commerce
(Autonomous)
Matunga, Mumbai – 400 019,
Maharashtra**

Syllabus for M.Sc. Semester II

Program: Master of Science

Course: Bioinformatics

Mandatory Paper I

**(As per Choice Based Semester and Grading System with effect
from Academic Year 2023-2024)**

M.Sc. Bioinformatics (2023-2024)

Mandatory Paper I

Course Title: Advance Biology & Structural Bioinformatics

Paper Code: GNKPSBI1502

Credits: 04

No. of Lectures: 60

Marks: 100

Learning Objectives:

- To teach concept of Central Dogma. Describe the general principles of gene organization and expression in both prokaryotic and eukaryotic organisms.
- To give a grounding in the Recombinant DNA technology and recent advancement in field of Molecular Biology.
- To study introduction of structural bioinformatics, conformational analysis of proteins and nucleic acids.
- To understand the protein structure prediction, and molecular interactions.
- This course will enable the students to: understand the levels of structural organization. of macromolecules and experimental methods of structure determination.
- Helps to understand the approaches for structure analysis and acquire knowledge of various algorithms & methods of structure prediction.
- Understand the genomic information.
- Describe the genome annotation pipeline, including manual and automated annotation

Learning Outcomes:

- Students will analyze structural-functional relationships of genes and proteins from bacteria to eukaryotes using genomic methods based on evolutionary relationships.
- Students will be able to explain the synthesis of proteins and nucleic acids with their regulation at the epigenetic, transcriptional, translational, and post-translational levels
- Acquire skills to use different approaches for prediction of protein structure.
- To understand the structural organization, structural properties and various techniques employed in the structure determination of Biological macromolecules – DNA, Protein.
- Advances understanding about primary and secondary structures of proteins
- Expansions understanding about tertiary and quaternary structures of proteins
- Designate about protein folding and its significance and helps to gain knowledge for the same.
- Use the UCSC genome browser to explore the available annotation tracks for a gene of interest.
- Helps in interpretation of the structure for a protein sequence based on an identified template.

Semester I

Unit		Title	No. of lectures
Unit 1		Concept of Central Dogma	
	1.1	Concept of Central Dogma & Recombinant DNA Technology	15
	1.2	DNA Replication, various enzymes involved, DNA topology, super coiling of DNA, origin of replication	
	1.3	Transcription – enzymes involved, sigma factor, DNA binding sites, initiation, elongation and termination	
	1.4	Translation or protein synthesis – enzymes or factors involved polyribosome	
	1.5	Operon model in prokaryotes and eukaryotes, lac operon, trp operon	
	1.6	Gene splicing, post translation modifications	
Unit 2		Recombinant DNA Technology	
	2.1	History of rDNA Technology, enzymes involved – restriction endonucleases, ligases etc.	15
	2.2	Cloning vectors – plasmids, bacteriophage vectors for eukaryotes- YAC.	
	2.3	DNA manipulation–cutting and joining DNA using nucleases and ligases, linkers and adapters	
	2.4	Cloning strategies, construction of libraries	
	2.5	Transformation of DNA into host cells, Screening for recombinants, blue white screening	
	2.6	DNA fingerprinting techniques	
	2.7	Application of recombinant DNA technology	
Unit 3		Molecular Biology Techniques	
	3.1	Protein/DNA/RNA probes, southern and western blotting, In situ hybridization	15
	3.2	DNA Sequencing	
	3.3	Microarray	
	3.4	Flow cytometry	
	3.5	Capillary Electrophoresis	
	3.6	PCR and Real time PCR	
Unit 4		Introduction to Protein Structures & Prediction Methods	
	4.1	Overview of structural bioinformatics – understanding structural basis for biological phenomena– challenges in structural bioinformatics	

			15
	4.2	Macromolecular Structure: Protein - Primary, Secondary, Super secondary, Tertiary and Quaternary structure, Potential energy maps, Ramachandran map, Co-ordinate systems	
	4.3	Macromolecular interactions: Protein – Protein (Proteomic analysis of Post-translational Modification- Phosphorylation, ubiquitination, and glycosylation) , Protein – Nucleic acids , Protein – carbohydrates, etc. Yeast two Hybrid Technique, Tools to study protein interaction: STRING, MIPS (other equivalent methods)	
	4.4	Principles of protein folding and methods to study protein folding, Anfinsen Experiment	
	4.5	Prediction of protein structure: secondary structure prediction methods: First, second and third generation methods.	
	4.6	Prediction of protein Tertiary structure prediction: Homology modeling, fold recognition and ab initio methods	
	4.7	Validation of protein tertiary structure using Ramachandran Plot/RAMPAGE/SAVES (or other equivalent methods)	

Semester II Evaluation Pattern

Internal:25

External: 75

M.Sc. Bioinformatics (2023-2024)

Mandatory Paper I - Practical

Course Title: Advance Biology & Bioinformatics

Paper Code: GNKPSBI1P502

Credits: 02

Marks: 50

1. Isolation of plasmid DNA and chromosomal DNA
2. Restriction digestion of DNA
3. Ligation
4. PAGE separation of human serum proteins
5. To perform Polymerase Chain Reaction of gene of interest
6. Demonstration of DNA Sequencer
7. Demonstration of Flow cytometer
8. Demonstration of Real time PCR
9. Secondary structure prediction using various available tools
10. Tertiary structure and function prediction using homology modeling, Fold recognition and ab initio method
 - a. Modeller
 - b. I TASSAR

- c. ROBETTA / ROSETTA
11. Validation of predicted structure by SAVES server or verify 3d (other equivalent methods)
12. Visualization of 3D protein structure using
 - a. Rasmol
 - b. PyMol
 - c. Chimera
13. Binding pocket prediction – Castp; Glycosylation, phosphorylation sites prediction (other equivalent methods)
14. Structural Blast – VAST & DALI

References:

- Primrose, S.B. and Twyman, R.M. (2006) Principles of Genetic Manipulation and Genomics. Seventh Edition. Blackwell Publishing, USA.
- Winnacker, E-L.(1987) From Genes to Clones. VCH Publishers, USA.
- Sambrook J. and Russell D. 2001. Molecular Cloning: A Laboratory Manual, 3rd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- B. M. Turner, Chromatin and Gene Regulation: Molecular Mechanisms in Epigenetics
- Russell Peter J. iGenetics: A Molecular Approach 3rd ed.: Pearson International Edition. Publisher: New York, Pearson ISBN: 9780321610225.
- Cesareni Giovanni, Gimona Mario, Sudol Marius, Yaffe Michael (Editors). Modular Protein Domains. Publisher: Weinheim Wiley-VCH. 2005. ISBN: 352730813X.
- Höltje Hans-Dieter, Sippl Wolfgang, Rognan Didier, Folkers Gerd. Molecular Modeling: Basic Principles and Applications. Publisher: New York, Wiley-VCH. 2003. ISBN: 3527305890.
- Webster David (Editor). Protein Structure Prediction: Methods and Protocols (Methods in Molecular Biology) Volume 143. Publisher: New Jersey Humana Press. 2000. ISBN: 0896036375.
- Sternberg Michael J. E. Protein Structure Prediction: A Practical Approach. Publisher: USA, Oxford University Press. 1997. ISBN: 0199634953.
- R.M. Twyman, “Principles of Proteomics (Advanced Text Series)”, Bios Scientific, 2004. 2
- Daniel C. Liebler, ” Introduction to Proteomics: Tools for the New Biology”, Humana Press Inc., 2002
- Ian M. Rosenberg, “Protein Analysis and Purification: Benchtop Techniques “- , Springer, 2005 2. Philip L.R. Bonner, “Protein Purification”, Taylor & Francis, 2007
- David W Mount, “Bioinformatics- Sequence and genome analysis”, Cold Spring Harbor Laboratory Press, second edition, 2004
- S. R. Pennington, M. J. Dunn, “Proteomics: from Protein Sequence to Function”, Springer publications, first edition, 2001.
- Timothy Palzkill, “Proteomics”, Springer, 2002.
- Igor Jurisica, Dennis Wile. Knowledge Discovery in Proteomics. 2006. Chapman & Hall / CRC Pennington SR (Ed), Dunn M. J. (Ed) Proteomics: from protein sequence to function. 2002 Viva Books Pvt. Ltd.
- Srivastava Sudhir (Ed). Informatics in Proteomics 2005 Taylor & Francis Group / CRC



Shiromani Gurdwara Parbandhak Committee's

**Guru Nanak Khalsa College of Arts, Science and Commerce
(Autonomous)
Matunga, Mumbai – 400 019, Maharashtra**

Syllabus for M.Sc. Semester II

Program: Master of Science

Course: Bioinformatics

Mandatory Paper: II

**(As per Choice Based Semester and Grading System with effect from
Academic Year 2023-2024)**

M.Sc. Bioinformatics (2023-2024)

Mandatory Paper II

Course Title: Java Programming, Introduction to Linux and Machine Learning

Paper Code: GNKPSBI2502

Credits: 04

No. of Lectures:

Marks: 100

Learning Objectives:

- Gain knowledge about basic Java language syntax and semantics to write Java programs and use concepts such as variables, conditional and iterative execution methods etc.
- Understand the fundamentals of object-oriented programming in Java, including defining classes, objects, invoking methods etc. and exception handling mechanisms.
- Understand the principles of inheritance, packages and interfaces.
- Machine Learning is broad and fast growing sub-field of Artificial Intelligence.
- This course introduces students to the basic concepts and techniques of Machine Learning.
- The objective of this course is to develop the skills required for Machine Learning Technologies with use of Python to analyze data, create beautiful visualizations, and problem solving using powerful machine learning algorithms.

Learning Outcomes:

- Identify classes, objects, members of a class and relationships among them needed for a specific problem.
- Have a good understanding of the fundamental issues and challenges of machine learning: data, model selection, model complexity, etc.
- Appreciate the underlying mathematical relationships within and across Machine Learning algorithms and the paradigms of supervised and un-supervised learning.
- Be able to design and implement various machine learning algorithms in a range of real-world applications.
- Application of Biological data using ML.
- Implementation & Evaluation of various machine learning algorithms

Unit		Title	No. of lectures
Unit 1		Core Java and Advanced Java	15
	1.1	An introduction to Object-Oriented Programming	
	1.2	Object-oriented programming and Java	
	1.3	Introduction to Java basics	
	1.4	Objects, Arrays, Conditionals and Loops	
	1.5	Creating Classes and Applications in Java, methods	
	1.6	Managing Simple Events and Interactivity	

	1.7	Creating User Interfaces with AWT	
	1.8	Packages and Interfaces	
	1.9	Exception	
Unit 2		JDBC	
	2.1	Introduction to JDBC	
	2.2	Client Server Application	
	2.3	Java Drivers: Type 1 Drivers, Type 2 Drivers, Type 3 Drivers, Type 4 Drivers	15
	2.4	Difference Between Java Drivers	
	2.5	Executing SQL Statements	
	2.6	Programs on Open a JDBC connection to connect database, create a table using JDBC	
	2.7	Insert values using JDBC, retrieve data using JDBC, Close a JDBC connection	
Unit 3		Introduction to Linux and Basic Linux commands	
	3.1	History and design	
	3.2	Principles of Linux	
	3.3	Functions of Linux OS	15
	3.4	Basic shell commands	
	3.5	Understanding Linux file permissions	
	3.6	Basic script building and File creation in Linux	
Unit 4		Introduction to Machine Learning in Bioinformatics	
	4.1	Introduction	
	4.2	Importance of machine learning	
	4.3	Types of machine learning	15
	4.4	Supervised Learning	
	4.5	Unsupervised learning	
	4.6	Semi supervised learning	

Semester II Evaluation Pattern

Internal: 25

External: 75

M.Sc. Bioinformatics (2023-2024)

Mandatory Paper II - Practical

Course Title: Java Programming, Introduction to Linux and Machine Learning

Paper Code: GNKPSBI2P502

Credits: 02

Marks: 50

1. Programs on Classes and Constructors, Destructor
 - a. Create a class and functions
 - b. Create a class and constructor to allocate memory
 - c. Create destructor to deallocate the memory

2. Programs on Inheritance and Polymorphism:
 - a. Implement Single Inheritance
 - b. Implement Multiple Inheritance
 - c. Implement Multilevel Inheritance
 - d. Implement Hierarchical Inheritance
 - e. Create Runtime Polymorphism
 - f. Create Compile Time Polymorphism

3. Arrays and Loops and Conditional Statement
 - a. Create and access 1-D array
 - b. Create and access 2-D array
 - c. Implement while loop
 - d. Implement do-while loop
 - e. Implement for loop
 - f. Implement if conditional statement
 - g. Implement if-else conditional statement
 - h. Implement Nested if-else conditional statement

4. Create AWT components on:
 - a. Frame
 - b. Button
 - c. Text Field
 - d. Checkbox
 - e. Radio Button

5. Create SWING components on:
 - a. JFrame
 - b. JButton
 - c. JText Field
 - d. JCheckbox
 - e. JRadio Button
 - f. JOptionPane
6. Create programs on packages
7. JDBC
 - a. Open a JDBC connection
 - b. Create a table using JDBC
 - c. Insert values using JDBC
 - d. Retrieve data using JDBC
 - e. Close a JDBC connection
8. LINUX Commands
 - a. Demonstration of Linux installment
 - b. Understand the working of UNIX, UNIX file system
 - c. Basic UNIX commands
 - File System Commands
 - ls and ls with options
 - d. Study of Commands:
 - tr, head, tail, wc, sort
 - find, file, ,split
 - e. Comparing files: od, cmp, comm, diff, uniq
 - f. Filter Commands: Grep, egrep, fgrep

References:

1. Java: The Complete Reference (English, Schildt Herbert).
2. Core and Advanced Java, Black Book, Recommended by CDAC, Revised and Upgraded Kindle Edition
3. Fundamentals of Linux: Explore the essentials of the Linux command line Kindle Edition
4. Practical Guide to Linux Commands Editor, by Mark G. Sobell.
5. Breiman, Leo, Friedman, J. H., Olshen, R. A. and Stone, C. J. (1984).
6. Classification and regression trees. Monterey, CA: Wadsworth and Brooks/Cole Advanced Books and Software.



**Shiromani Gurdwara Parbandhak Committee's
Guru Nanak Khalsa College of Arts, Science and Commerce
(Autonomous)
Matunga, Mumbai – 400 019, Maharashtra**

Syllabus for M.Sc. Semester II

Program: Master of Science

Course: Bioinformatics

ELECTIVE PAPER

**(As per Choice Based Semester and Grading System with effect from
Academic Year 2023-2024)**

M.Sc. Bioinformatics (2023- 2024)

Elective Paper

Course Title: Genomics & Proteomics in Bioinformatics, NGS

Paper Code: GNKPSBI3502

Credits: 03

No. of Lectures: 45

Marks: 100

Learning Objectives:

- To study prokaryotic and eukaryotic genomes, general methods of genome sequencing techniques, genome analysis and annotations, genome mapping techniques and applications of genomics.
- To understand the proteins enclosed by the genes with respect to structure, function, protein – protein interactions, techniques for separation and analysis, database and applications.
- Thorough knowledge of basic concepts underpinning Bioinformatics & NGS Data
- Define whole genome sequencing.
- Knowledge and skill in NGS Data Analytics
- Appreciate and explore the comparative genomics databases.
- Recognize various algorithms used for comparisons of full genome and gene order.
- Know the concepts of SNPs and their significance.
- Realize and appreciate the proteomics concepts and technology.

Learning Outcomes:

- Appreciate the importance of full genome comparisons and helps in further research studies.
- Genome mapping solves the complicated puzzle of genes with pieces of information coming from laboratories all over the world and helps to detect the functions of the genes.
- Helps of proteomics studies may lead to better treatments in future.
- Understand the different platforms for first and second-generation sequencing
- Understand the advantage and disadvantage of different types of first and second generation sequencing
- Understand the different platforms for third generation sequencing.
- Understand the different types of high throughput sequencing.
- Understand the library preparation methods for different types of high throughput sequencing.
- Understand the types of sequencing methods used in different platforms
- Understand the different file formats used in NGS data analysis
- Understand the different sources of publicly available NGS datasets and how to retrieve them
-

Unit		Title	No. of lectures
Unit 1		Genomics & Proteomics	15
	1.1	Introduction to Genomics & Proteomics, Need, scope and challenges, Applications	
	1.2	Prediction of Introns, Exons and its identification tools, Open Reading Frame (ORFs)	
	1.3	Predictions of genes, promoters, splice sites, regulatory regions: basic principles, application of methods to prokaryotic and eukaryotic genomes and interpretation of results. (Synteny and gene order) Mutations and diseases	
	1.4	Identification of disease genes using OMIM database (Case study), Microarray studies, reference genome sequence, integrated genomic maps, gene expression profiling	
	1.5	Orphan GPCRs, Identification of Orphan GPCRs, Reverse Pharmacology, Reverse Pharmacology for Orphan GPCRs, Deorphanisation	
	1.6	Protein Identification with antibodies, protein sequence determination by chemical degradation/ Edman's degradation, Short-gun proteomics for proteome profile	
	1.7	Computational tools- advanced tools for data analysis at ExPasy, INTERPRO and other equivalent tools	
Unit 2		History & evolution of sequencing	15
	2.1	GENE prediction & Genome Alignments: BLAST2GO, MUMmer, Pip Maker, VISTA, Comparison of Gene Order Comparative Genomics Databases: COG, VirGen, CORG, Single Nucleotide Polymorphism, dbSNP and other SNP-related databases	
	2.2	History & evolution of sequencing: First-generation technologies – Sanger dideoxy sequencing, Maxam-Gilbert sequencing. Technologies used in Human Genome Project. Shotgun sequencing. Advantages and disadvantages of first generation sequencing. Next (second)-generation sequencing – difference between first and next generation sequencing, NGS platforms – Roche 454, ABI SOLiD, Ion torrent, Illumina. Advantages and disadvantages of second generation sequencing.	

	2.3	Third-generation sequencing: PacBio, Oxford nanopore. Advantages and disadvantages of third generation sequencing. Comparison of output, accuracy and types of errors of first, second and third generation sequencing technologies. Introduction to NGS technologies: DNA-seq, RNA-seq, ChIP-seq, Hi-C, Metagenomics, Single cell sequencing.	
Unit 3		Data Analysis in Genomics and Transcriptomics	
	3.1	Preparation for sequencing: Different sample preparation methods for different type of NGS (DNaseq, RNAseq, ChIPseq, Metagenomics, Single cell), Adaptors, Index, Barcode. Library preparation methods - Bridge amplification, Emulsion PCR. Sequencing methods – sequencing by synthesis, ion semiconductor, SMRT, nanopore.	
	3.2	Data formats & Pre-processing: Data formats overview – FASTQ, subreads, nanopore data, single cell data. Single-end, Paired-end, Mate-pair. NGS Data sources – NCBI , SRA, EBI-ENA, DDBJ-SRA, GEO; Retrieving data from data sources - SRA toolkit; Aspera connect. Sequence quality measures – Phred quality score. Quality check – tool – FASTQC, Pre-processing: Trimmomatic, Fastx-toolkit.	15
	3.3	Introduction to NGS Data Analysis: Assembly (overview) -principles, output file formats – contigs, scaffolds, assembly quality assessment – N50, total length, no. of contigs/scaffolds; Mapping (overview) – Principles, tools – BWA, Bowtie, SAMtools, output file formats – BAM, SAM, mapping alignment assessment – no. of reads mapped, concordant reads; Visualizations tools – IGV.	

Semester II Evaluation Pattern

Internal: 25

External: 75

M.Sc. Bioinformatics (2023-2024)

Elective Paper Practical

Course Title: Genomics & Proteomics in Bioinformatics, NGS

Paper Code: GNKPSBI3P502

Credits: 01

Marks: 50

1. Detection of general signals, regulatory regions and promoter prediction (Minimum 5)
2. Microarray Studies: GEO database
3. GPCR Database
4. EST Database
5. Functional annotation and analysis of genome-scale sequence: BLAST2GO
6. SNP Database: DbSNP & identification of mutational genes
7. Whole Genome Sequencing:
 - a. MUMmer
 - b. Pipmaker
 - c. VISTA
8. Identification of repetitive elements: Repeat Masker Tool
9. NGS:
 - a. Working with NGS databases and file formats
 - b. Variant calling and filtering using BCF tools
 - c. SAMtools
 - d. FASTQC
 - e. Cutadapt & Trimmomatic
 - f. Reference genome mapping and alignment using BWA /bowtie
 - g. De Novo assembly using velvet/spades
 - h. Determining quality of the assembly using Quast tool
 - i. Metagenomics
 - j. Prokka
 - k. Visualize Insertions/deletion

References:

1. Sándor Suhai, Genomics and Proteomics: Functional and Computational Aspects, Kluwer Academic Publishers, 2002
2. Azuaje F & Dopazo J. 2005. Data Analysis and Visualization in Genomics and Proteomics. John Wiley & Sons.
3. Akay M. (Ed) Genomics and Proteomics Engineering in Medicine and Biology 2007 WileyInterscience John Wiley & sons, Inc. Publication, USA
4. “Microbial Genomics And Proteomics” by Niyaz Ahmed
5. Christendat D, et al. Structural proteomics of an archeon. Nat. Struct. Biol. 2000;7(10):903–909. [PubMed]
6. Discovering Genomics, Proteomics and Bioinformatics 2nd edition - by A. Malcolm Campbell and Laurie J. Heyer. by Cold Spring Harbor Laboratory Press 2006.
7. Principles of Genome Analysis and Genomics (3rd Ed.) by Primrose, S.B. and Twyman, R.M., Blackwell Publishing Company, Oxford, UK. 2003
8. Introduction to Proteomics – Tools for the new biology (1st Ed.) by Liebler, D.C., Humana Press Inc., New Jersey, USA. 2002
9. Bioinformatics and Functional Genomics by Pevsner, J., John Wiley and Sons, New Jersey, USA. 2003
10. Bioinformatics: Sequence and Genome Analysis by Mount, D., Cold Spring Harbor Laboratory Press, New York. 2004